Cell

Science Networking Gets Serious

A new breed of networking applications offers scientists much more than typical social networking sites, but how useful are they? Amy Maxmen reports.

Social networking sites might significantly advance collaborative research, if only more scientists liked them. Although Web 2.0 offers infinite possibilities for forming online collaborations, traditional social networking sites like Facebook and Twitter and professional networking sites like ResearchGATE and LinkedIn seem to be underused by the scientific community. However, new applications are emerging that help busy researchers with their work, while unobtrusively connecting them through the data they enter online, be it bookmarking a paper or annotating a gene.

Biochemist Cameron Neylon at the University of Southampton, UK, who blogs about topics like online lab notebooks, online collaborations, and social networks for scientists, says, "I think the reason why the social networks out there have failed to capture the interest of most scientists is that they solve a problem that most researchers don't think they have. They are based on the presumption that scientists need to find other scientists, whereas most scientists don't feel they have a real problem finding people to connect with. When they want to solve a specific problem, they can simply go to the literature and find an author on the web." Instead, he says, "Tools that help scientists spend more time at the bench, and allow them to get more out of the time they spend at the computer, will succeed where most of the peripheral distracting networking tools have failed."

Grant funding agencies are becoming interested in social networking applications that can help researchers. For example, last November, the National Institutes of Health (NIH) awarded $27 million in Recovery Act (ARRA) funds to two research teams (http://www.nih.gov/news/health/nov2009/ncrr-02.htm). One team, led by researchers at the University of Florida, is developing a social networking site for scientists called VIVO. And the other team, led by Harvard University Medical School investigators, is developing a networking site named eagle-i to connect scientists to resources like reagents or cell lines located at other institutions. NIH program officer Elaine Collier explains, "There was an opportunity with the ARRA funds to award innovative ideas in biomedical research that would move the field forward, could be done in two years, and could employ a lot of people…Ultimately, our vision is that there will be a national network to find resources and scientists."

However, there is skepticism about whether traditional social networking sites will ever be actively used by scientists if the sites don't provide some other immediate function. "The problem I see with the recent networks funded by the NIH is that they don't seem to be driven by the needs and the problems people have immediately," says Victor Henning, the founder and director of Mendeley (http://www.mendeley.com), a science literature management tool with features that enable networking. "Sure, Facebook took off when it started at Harvard because it indexed people with photos, and therefore, it could help you find out who that hot girl in your class was. With research, it's a different scenario. It's not hard to find most faculty on the web through a lab homepage or through the literature. What people want is a way to handle accumulating data and ways to manage their workflow."

With Mendeley, Henning hasn't tossed out the concept of social networking for scientists, but rather he's incorporated it into something larger. "We wanted to create software that would have value in itself. So although Mendeley doesn't depend on network effects, there is a social layer on top so that as a result, each document on the site comes with anonymized demographic information about who reads it, what university they're at, and what their academic status is," says Henning. Unlike sites created to digitize connections between people, Mendeley follows in the footsteps of applications like Last.fm, a popular online radio site that uses social networking software to aggregate people based on commonalities in their listening habits. Henning says, "In traditional networks, like Facebook, you look for friends. The difference in Mendeley is that it's first about connecting data—which articles are related, what those relationships are, and who reads them—and then you let the social network emerge based on those connections."

Getting Connected in the Cloud

With more than 40 million monthly users, Last.fm provided Henning with a successful model to follow. Because Last.fm operates in the so-called internet "cloud" looming above personal computers, it has the computing space and power needed to analyze vast amounts of data about the music people play. Henning wanted to emulate Last.fm using research papers rather than music as currency. Mendeley helps investigators to index and organize the pdfs they comb through. And Mendeley's social networking features help scientists discover new papers by revealing what the most popular papers are each week and which scientists read them.

Mendeley's online predecessor, CiteULike (http://www.citeulike.org), helps scientists to discover new references through social networking features as well. By clicking on manuscripts online, researchers add citations to a running list kept on the web that is searchable by tagged words. Because the lists are public by default, users can see who maintains libraries like theirs. And with "watchlists," they can track those users to find out when they add a reference. Unlike traditional social networking applications designed to connect people from the outset, these sites operate on a model best described by Microsoft software developer Jon Udell, who says, "Data finds people, then people find people."

The rise of "cloud computing" should be a major asset to scientists who analyze "exafloods" of data transmitted over the Internet. By using a third party to host data, cloud-based software can generally analyze information faster and for less money than projects run on local computer clusters. Investigators working in bioinformatics, genomics, and astrophysics have been among the first to embrace cloud computing because of the huge datasets generated and the computational power needed to analyze them. Likewise, applications that store loads of data and run the statistics required for social networking features tend to operate in the cloud as well. "I'm making the assumption that latent social networks will begin to launch as the use of cloud computing increases, because the more data you have on the web, the more you can start looking for these hidden connections," Henning says.

Seeing the potential for how cloud computing can benefit researchers, the National Science Foundation (NSF) announced in February this year that it would team up with Microsoft to offer free access to cloud computing resources for 3 years to researchers with winning proposals presenting clever ways to exploit cloud capabilities. Jeannette Wing, the head of Computer and Information Science and Engineering at the NSF, is encouraging investigators to team up with computer scientists when developing their ideas. She says, "Once data sets get large, investigators can't just ftp each other data. And it's also not clear what institutions are able to host a lot of data, so putting it all in the cloud is a trend that's going to happen."

Harnessing Community Intelligence

Gene annotation and other curatorial sites already take a collaborative approach by relying on contributions from the scientific community, and with the right software they can connect scientists through their activity online. With his gene annotation project, BioGPS, Andrew Su, associate director of bioinformatics at the Genomics Institute of the Novartis Research Foundation in San Diego, California, is testing out the theory that latent communities that form within an application as people contribute to the site make the site more useful over time. "A big question now is how to engage communities of scientists to collaboratively work together," Su says. "The internet enables us to collaboratively do science, and I'm operating on the idea that networks of people can be useful and productive." In BioGPS, investigators decide what gene-related content they'd like to see on a webpage. Su compares it to customizable iGoogle homepages, except that instead of opting to see updates on weather, celebrities, and the local news, users choose to see gene lists, 3D crystal structures, and gene expression data. They can also import small computer programs, called plugins, when the information they'd like to see isn't already in BioGPS. Of the 200 or so plugins added to BioGPS since its launch in 2008, several reveal molecular biology reagents available for a gene of interest, others show biological pathways relevant to a gene, and still others provide alternative splicing information. As users customize their profiles they automatically feed information into the system, which enables latent connections to form. With social networking features, BioGPS computes what the most popular plugins are, and soon, it will let users know who maintains a gene list similar to theirs, unless the user requests privacy. "Discovered similarities with other users could in some cases be more valuable than the original reason a user came to BioGPS, which was to view gene information provided by the plugins," Su says.

In order to encourage researcher participation in Rfam, a database of RNA annotations and alignments, Alex Bateman, a computational biologist at the Wellcome Trust Sanger Institute near Cambridge, UK, lures researchers with publication incentives. Wikipedia automatically feeds annotation information into Rfam. And because Rfam benefits by having data first entered into Wikipedia where Wikipedia's online community edits the entry, Bateman has convinced the editors at the journal *RNA Biology* to add a requirement for publication of RNA families. In addition to submitting a manuscript to the journal, authors must write or update a short Wikipedia page that summarizes their work. As a result, researchers contribute as a means to publish but end up collaborating on an online project. And they tend to keep an eye on their pages long after publication, says Bateman, who monitors the back-and-forth edits that occur on RNA pages over time. "I wouldn't say I was a big believer in Web 2.0. But I was taken on board as it delivered useful things. I must say that social networking and the community style of Web 2.0 has been fantastic for enabling community annotation," he says.

David Lipman, director of the National Center for Biotechnology Information (NCBI), comments, "The funny thing about Rfam is that you'd think there would be many more databases like this. Everybody who does a curated database in biology thinks about how we can leverage Wikipedia for scientific curation." GenBank, one of NCBI's great online successes, has relied on community input since the late eighties. But unlike Wikipedia, the authors' original contributions remain more or less frozen in time. Wikipedia, and related "Wiki" websites, evolve as various contributors add and edit them over time. Networks of people collaborate to keep these sites up-to-date. And although the information uploaded is not always accurate, volunteer editors have been shown to weed out most inaccuracies quickly. Science-related Wiki pages are multiplying. For example, researchers add and edit protocols and other laboratory-related information on the Wiki OpenWetWare; biologists catalog gene functions on WikiGenes; and they enter data on single-nucleotide polymorphisms (SNPs) into SNPedia. NIH immunologist Ron Germain can easily imagine how a Wikipedia-like website that counted on community participation could help systems biologists. "It would be great to have one place where systems biology people could share their models, download, use, and add to existing models. I don't care if it's pre- or post-publication, it's just a matter of having a simple way to share things over the Web." The lure of sites like these is that a researcher may be more apt to participate because the site directly contributes to their research.

Lipman questions why universities haven't pounced on Wikipedia and Wiki pages. "What I wonder is why professors don't curate [pages on] Wikipedia and add course materials and open

access sections of textbooks, much of which they post online anyways. We aren't really seeing the potential that you would hope for with all of the Web 2.0 tools out there. We aren't seeing the academic community take advantage of them as much as other subsets of the community." He adds, "I feel that if you really wanted to see these things move more quickly, it's not the funding agencies that have to lead but the universities. If they don't, the intellectual center of mass will move off of the university campus and onto the Web and into the cloud. Academics should be using these tools already. And, in fact, there could be a huge risk if universities don't get more involved in social networking and in Web 2.0 to create a center of gravity for education and research. You could imagine the reputation of a university being incredibly enhanced if more people used resources that might easily be posted online. The academic response to the potential out there has not been that inspired."

So will these new applications that subtly bring together networks of researchers for reasons other than social networking entice more scientists? "The thing about social networks is, that for all intents and purposes, this is only about five years old. And so everything feels new and shiny, but these things are actually very immature," says biochemist and blogger Neylon. "So what we should really be thinking about is what we might want in the future."

Amy Maxmen
New York, NY, USA
DOI 10.1016/j.cell.2010.04.019